

# METODYKA PRACY Z SĘDZIAMI KOMPETENTNYMI W PROCESIE OPRACOWYWANIA SKALI DO POMIARU KONSTRUKTÓW: REKOMENDACJE I EGZEMPLIFIKACJA

DOI: 10.33141/po.2020.12.02

Przegląd Organizacji, Nr 12(971), 2020, s. 13-18

www.przegladorganizacji.pl

Maciej Czarnecki

© Towarzystwo Naukowe Organizacji i Kierownictwa (TNOiK)

## Wprowadzenie

Jednym z obszarów zainteresowań nauk o zarządzaniu jest odkrywanie zależności pomiędzy zjawiskami zachodzącymi w organizacjach. Ich potwierdzanie w badaniach ilościowych wymaga formułowania hipotez opisujących zależności między zmiennymi. Wiele z nich nie może być zaobserwowanych w sposób bezpośredni. Takie zmienne nazywane są zmiennymi ukrytymi. W celu zbadania zmiennej ukrytej należy dokonać jej operacjonalizacji, co w badaniach ankietowych oznacza uchwycenie jej istoty za pomocą co najmniej jednego pytania (Lavrakas, 2008, s. 939). Jakość skal służących do pomiaru zjawisk będących przedmiotem zainteresowania badacza jest punktem wyjścia w drodze do uzyskania wiarygodnych wyników. Rezultatem stosowania narzędzi opracowanych w sposób niespełniający zasad rygoru metodologicznego są poważne wątpliwości dotyczące rzetelności ustaleń i ich wiarygodności (Czakon, 2019, s. 3). Pomimo tego poważne błędy w metodyce opracowywania narzędzi pomiarowych, często nawet jej całkowite ignorowanie, to jeden z najczęstszych zarzutów dotyczących badań w naukach o zarządzaniu (Govindarajan, Kopal, 2006, s. 190) oraz jedna z głównych barier ich rozwoju (Danneels, 2016, s. 2183). Zbiór powszechnie uznanych narzędzi pomiarowych jest bardzo mały (Czakon 2019, s. 8), a do wyjątków należą konstrukty, które doczekały się jednej bądź kilku uznanych skal pomiarowych. Warto wskazać, iż korzystanie z pomyślnie zwalidowanych<sup>1</sup> wcześniej skal opracowanych przez innych badaczy, pozwoliłoby na replikację zarówno skal, jak i badań, a więc zwiększyłoby szanse na porównywanie uzyskiwanych wyników. Duża część badaczy przedstawia jednak własne, nowe skale pomiarowe. W konsekwencji znacząca część konstruktywów powszechnie używanych w literaturze dotyczącej zarządzania nie doczekała się takich miar (np. ciągłe doskonalenie, samoorganizacja itd.). Jedną z przyczyn tego stanu rzeczy może być wysoki stopień trudności metodyki mającej na celu opracowanie takiego narzędzia, kolejną zaś – niedostateczna ilość wskazówek dotyczących jednego z początkowych etapów procedury badawczej, jakim jest selekcja wstępnych pozycji kwestionariuszowych<sup>2</sup> do ankiety badań pilotażowych z udziałem tzw. *sędziów kompetentnych*.

Niniejszy artykuł stanowi próbę wypełnienia wskazanej luki. Jego celem jest przedstawienie procedury pracy z sędziami kompetentnymi przy selekcji wstępnych pozycji kwestionariuszowych do ankiety badań pilotażowych oraz jej egzemplifikacja. W treści artykułu przedstawiono krótki przegląd literatury, wskazując na dylematy, przed którymi stoi badacz, korzystając z pomocy zewnętrznych ekspertów, następnie przedstawiono przykład postępowania w tym zakresie, będący częścią procedury badawczej w zrealizowanych przez autora badaniach związanych z opracowaniem nowych skal.

## Badanie opinii sędziów kompetentnych w procesie opracowywania narzędzia pomiarowego

Procedura opracowywania narzędzia pomiarowego zaproponowana przez G.A. Churchilla (1979), będąca najczęściej stosowaną klasyczną procedurą (Czakon, 2019, s. 8), obejmuje między innymi skorzystanie z pomocy ekspertów zewnętrznych. Na podstawie przeglądu literatury i/lub badań jakościowych badacz opracowuje listę wstępnych pytań do ankiety. Samodzielna kwalifikacja części z nich do kwestionariusza badania pilotażowego oparta byłaby na subiektywnym odczuciu i doświadczeniu badacza, mogłaby więc prowadzić do wystąpienia błędu selekcji. W celu jego uniknięcia w proces selekcji należy zaangażować ekspertów, którzy za pomocą przyjętej skali ocenią każdą ze wstępnych pozycji. Sędziów warto poprosić także o komentarze dotyczące ocenianych stwierdzeń, np. czy pasują do definicji konstruktów, czy są jednoznacznie sformułowane, czy stylistyka sformułowań jest poprawna itd. (DeVellis, 2017, s. 128).

Wielu badaczy opracowujących nowe skale pomiarowe zaniedbuje jednak ten etap badań (Hardesty, Bearden, 2004, s. 99). Co więcej, tylko część prac przedstawiających proces ich opracowywania zawiera opis zastosowanej procedury pracy z sędziami. Większość autorów poprzestaje na zamieszczeniu jedynie ogólnikowych informacji w tym zakresie (Gorbaniuk, 2016, s. 6), ograniczając się do wskazania liczby zaangażowanych sędziów bądź też czy byli oni



menedżerami czy badaczami zajmującymi się daną tematyką itp. Jedną z przyczyn takiego stanu rzeczy mogą być niejasności związane z metodyką badania opinii sędziów.

Pewne wątpliwości mogą pojawić się już na etapie decyzji dotyczącej doboru sędziów. W literaturze przedmiotu trudno znaleźć wytyczne dotyczące tej kwestii, poza dość ogólnymi wskazówkami, iż należy do tej roli zaangażować ekspertów mających doświadczenie w pracy badawczej w danym obszarze i/lub praktyków z adekwatnym przygotowaniem akademickim (Crawford, Kelder, 2019, s. 136). Dla przykładu, R.K. Rai (2016, s. 13), opracowując skalę tworzenia wartości w kontekście aliansów strategicznych, zaangażował w proces walidacji skali czterech praktyków oraz czterech naukowców. Zdarzają się jednak przypadki angażowania wyłącznie naukowców (Thomas, Lucas, 2019) czy nawet doktorantów (Liang i in., 2012, s. 78). Z dokonanego przez autora przeglądu literatury wynika, że bardziej powszechną praktyką jest angażowanie jednocześnie naukowców oraz praktyków w podobnej liczebności.

Kolejny dylemat dotyczy liczby zaangażowanych w badanie sędziów. Interesujący jest fakt, iż wielokrotnie wznawiane, klasyczne pozycje zwarte dotyczące opracowywania skal pomiarowych (DeVellis, 2017) czy też badań marketingowych (Churchill, 2002), nie zawierają wskazówek dotyczących ich liczby. D.M. Hardesty i W. Bearden (2004, s. 101–102) wskazują, że wśród przeanalizowanych przez nich 200 opracowań dotyczących najczęściej używanych narzędzi pomiarowych w badaniach marketingowych liczba sędziów, o ile została podana, wahała się pomiędzy 3 a 50. E. Hornowska (2017, s. 233) sugeruje liczbę sędziów pomiędzy 3 a 40, zaś J.A. Crawford i J.A. Kelder (2019, s. 136) rekomendują zaangażowanie od 9 do 30 ekspertów. Liczbę sędziów można powiązać z oczekiwanym poziomem rzetelności skali mierzonej za pomocą wskaźnika *proportional reduction in loss*<sup>3</sup>. R.T. Rust i B. Cooil (1994) wskazują, że przy zastosowaniu trzystopniowej skali Likerta i wskaźnika zgodności odpowiedzi ekspertów wynoszącym 0,5 zaangażowanie w badania dziewięciu sędziów zapewnia osiągnięcie tego wskaźnika na poziomie 0,91 (Rust, Cooil, 1994, s. 8), przy zastosowaniu skali czterostopniowej do osiągnięcia takiego poziomu wystarczy 7 sędziów (Rust, Cooil, 1994, s. 10), zaś skala pięciostopniowa wymaga zaangażowania tylko 6 sędziów (Rust, Cooil, 1994, s. 12).

Kolejne wątpliwości wiążą się z wyborem metody, za pomocą której sędziowie powinni ocenić wstępne pozycje kwestionariusza. Oceny tej mogą dokonać zgodnie z propozycją J.L. Zaichkowsky'ego (1985, s. 343), zgodnie z którą każde z pytań jest oceniane jako wyraźnie reprezentatywne/ w pewnym stopniu reprezentatywne lub niereprezentatywne dla konstruktów będącego przedmiotem zainteresowania. Podobnie E. Hornowska (2017, s. 87) proponuje, aby każda pozycja została oceniona za pomocą trzystopniowej skali: *pozycja ma zasadnicze znaczenie dla testu/ pozycja jest użyteczna, jednak nie ma zasadniczego znaczenia/ pozycja nie powinna znaleźć się w teście*. Z kolei C. Obermiller i E.R. Spangenberg (1988, s. 170) sugerują użycie czterostopniowej skali Likerta (*pozycja bardzo dobrze/ dobrze/ średnio/ słabo nadaje się do mierzenia konstruktów*). R.T. Rust i B. Cooil (1994, s. 9) przekonują, że użycie skal

o większej rozpiętości (np. pięciostopniowych zamiast trzystopniowych) pozwala na zaangażowanie mniejszej liczby sędziów przy uzyskiwaniu wyników o porównywalnej jakości, o czym wspomniano powyżej w kontekście liczby sędziów. Inną metodą wyrażania przez ekspertów opinii na temat pozycji kwestionariuszowych jest metoda rangowania, polegająca na sporządzaniu przez nich rankingów stwierdzeń (szeregowaniu od najlepszego do najgorszego bądź odwrotnie). Wykorzystana może być ona jednak wyłącznie w przypadku porównywania niewielkiej liczby pozycji (maksymalnie do 9 – Cabała, 2010, s. 49). Liczba ta będzie tym niższa, im wyższa była staranność badacza na etapie opracowywania listy wstępnych pozycji. W przypadku wysokiej jakości pracy na tym etapie sporządzenie rankingów wstępnych pozycji kwestionariusza sprawić będzie sędziom poważny problem, wszystkie one bowiem powinny dobrze nadawać się do pomiaru cechy (DeVellis, 2017, s. 128). Niektórzy badacze próbują poradzić sobie z tym problemem, przekształcając oceny wyrażone w skali punktowej w szeregi, istnieją jednak poważne wątpliwości, czy zabieg ten jest uprawniony (Cabała, 2010, s. 50).

Trudności przysparzać może także sposób analizy odpowiedzi sędziów. Nie ulega wątpliwości, że im zgodniej eksperci wskażą na przydatność danej pozycji do pomiaru cechy, z tym większym prawdopodobieństwem powinna ona zostać zakwalifikowana do kwestionariusza badań pilotażowych. Najprostszą metodą, biorącą pod uwagę zgodność opinii sędziów przy jednoczesnej selekcji pozycji, jest metoda Lawshego. Polega ona na przyjęciu do ankiety badań pilotażowych tych pozycji, które zostały wskazane przez większość sędziów jako istotne dla kwestionariusza (Hornowska, 2017, s. 233). Wielu badaczy kwalifikuje pozycję do ankiety badań pilotażowych w przypadku, gdy co najmniej 75% sędziów uzna ją za pożądaną (Hardesty, Bearden, 2004, s. 104). Warto zauważyć, że w przypadku użycia tej metody, badacz nie wie nawet w przybliżeniu, ile pozycji zostanie przez sędziów wyselekcjonowanych do dalszych etapów pracy nad kwestionariuszem. Wymusza ona także użycie trzystopniowej skali oceny stwierdzeń. Relatywnie prostym do zastosowania miernikiem oceny preferencji sędziów jest współczynnik zgodności uporządkowań wielokrotnych – wskaźnik konkordancji W Kendalla (Cabała, 2010, s. 37). Informuje on o stopniu zgodności opinii sędziów, nie wskazuje jednak pozycji, które należy przyjąć do ankiety badania pilotażowego. Ponadto można go stosować wyłącznie w przypadku użycia przez sędziów metody rangowania. O. Gorbaniuk (2016, s. 5) sugeruje, iż dobrą i prostą metodą analizy ocen sędziów jest posłużenie się uśrednionymi opiniami sędziów na temat każdej z pozycji kwestionariusza. Do kwestionariusza badań pilotażowych zakwalifikować należy kolejne stwierdzenia ze sporządzonego na tej podstawie rankingów.

Jak wynika z powyższych rozważań, badacz opracowujący nową skalę mierzy się z wieloma dylematami, wynikającymi z niedostatku wskazówek odnoszących się do metodyki pracy z sędziami kompetentnymi. Były one przedmiotem dyskusji, które zostały przeprowadzone przez autora z przedstawicielami nauk o zarządzaniu oraz ekspertami zajmującymi się statystyką<sup>4</sup> podczas pracy nad nowymi skalami do pomiaru dwóch konstruktów: organizacyjnego rozwoju oraz

organizacyjnych problemów wzrostu (Czarnecki, 2016). Poniżej zaprezentowana zostanie procedura pracy zastosowana przy opracowaniu tych skal, wypracowana z nimi w rezultacie tych dyskusji.

## Badanie opinii sędziów kompetentnych – egzemplifikacja

Wstępne pozycje kwestionariuszowe zostały opracowane na podstawie studiów literaturowych oraz wywiadów strukturyzowanych z dwudziestoma przedsiębiorcami i menedżerami. Zostały one następnie przedstawione dwóm sześciuosobowym grupom ekspertów zewnętrznych, składającym się z przedstawicieli nauki oraz praktyki. R.T. Rust i B. Cooil (1994, s. 12) stwierdzili, że dwunastu sędziów przy zgodności ich odpowiedzi wynoszącej 0,5 pozwala na uzyskanie wskaźnika rzetelności *proportional reduction in loss* na poziomie 0,98. Ekspertcy – praktycy byli menedżerami wysokiego szczebla pracującymi w przedsiębiorstwach średnich lub takich, które w okresie trzech lat poprzedzających wywiad były przedsiębiorstwami średnimi, lecz na skutek wzrostu stały się przedsiębiorstwami dużymi. Były to osoby odpowiedzialne m.in. za strategię firm, w tym także ich wzrost i rozwój. Ekspertcy – przedstawiciele nauk o zarządzaniu byli naukowcami posiadającymi stopień naukowy doktora habilitowanego, których przedmiotem zainteresowań były wzrost, rozwój i efektywność przedsiębiorstw. Przed podjęciem decyzji o zaangażowaniu w badanie w charakterze sędziów, autor przeprowadził rozmowy z każdym z kandydatów, w trakcie których zostali oni zapoznani z tematyką badań oraz oczekiwanymi efektami ich pracy. Rozmowy dały także okazję do sprawdzenia, czy wykazywali oni zainteresowanie udziałem w badaniu oraz czy ich kwalifikacje były wystarczające w kontekście jego celów. Ekspertcy zostali zapoznani z definicjami konstruktów oraz poproszeni o ocenę przydatności do ich pomiaru wstępnych pozycji kwestionariuszowych za pomocą pięciostopniowej skali Likerta. Sędziowie ocenili łącznie 195 wstępnych pozycji kwestionariuszowych. Redukcja ich liczby w następstwie przeprowadzenia badania opinii sędziów oraz badania pilotażowego została przedstawiona w tabeli 1.

Zgodnie z zaleceniami O. Gorbaniuka (2016), dla każdego stwierdzenia policzono średnie wewnętrzne (po odrzuceniu dwóch skrajnych ocen), następnie zaś na ich podstawie opracowano rankingi stwierdzeń. Do badania pilotażowego zakwalifikowano stwierdzenia, które uzyskały najwyższe średnie wewnętrzne. Pomocniczo korzystano z wartości odchyłeń standardowych odpowiedzi sędziów oraz porównań średnich wewnętrznych policzonych w obu grupach sędziów. Ich liczba wynosiła od 12 do 16 stwierdzeń dla różnych wymiarów rozwoju organizacyjnego oraz 32 dla organizacyjnych problemów wzrostu. Tabela 2 przedstawia wyniki oceny dokonanej przez sędziów dla rozwoju struktur organizacyjnych (jednego z pięciu wymiarów rozwoju organizacyjnego), prezentując średnie odchylenia standardowe, średnie wewnętrzne, jak również pozycje poszczególnych stwierdzeń w rankingu. Tabela podzielona została na trzy obszary. Lewa strona tabeli zawiera wskazane miary policzone dla odpowiedzi wszystkich sędziów (S). Obszary: środkowy oraz po prawej stronie, zawierają oceny poszczególnych stwierdzeń przyznane przez odpowiednio: sędziów – przedstawicieli nauki oraz sędziów – praktyków. Kolorem szarym oznaczono pytania, które zastosowano w badaniu pilotażowym, kolorem ciemnoszarym oznaczono te spośród nich, które następnie zastosowano w badaniu właściwym.

Sędziowie uznali część wstępnych pozycji kwestionariuszowych zaczerpniętych ze skal opublikowanych w literaturze, za słabo lub przeciętnie nadające się do pomiaru cech<sup>5</sup>. Sytuacja spowodowała dylemat: czy zaufać opinii sędziów, czy jednak uwzględnić w kwestionariuszu badań pilotażowych pozycje sugerowane w recenzowanych publikacjach. Podążanie za wynikami badania opinii sędziów, w przypadku selekcji wstępnych pozycji do kwestionariusza badań pilotażowych, nie jest bezwzględnym obowiązkiem badacza. Nie należy ich ignorować, jednak badacz powinien dokonać świadomej decyzji, jak je wykorzystać (DeVellis, 2017, s. 128). W opisywanych badaniach zdecydowano o uwzględnieniu w całości opinii sędziów, w konsekwencji zaś nie włączono nisko i przeciętnie ocenionych przez nich pozycji do kwestionariusza badania pilotażowego. Sędziowie okazali się także pomocni w podjęciu decyzji dotyczącej sposobu mierzenia dynamiki zmiennych. Rozstrzygnęli

Tabela 1. Redukcja liczby pozycji skal w poszczególnych etapach ich tworzenia

Konstrukt / wymiar konstruktów	Liczba pozycji przedłożonych ocenie sędziom kompetentnym	Liczba pozycji w badaniu pilotażowym	Liczba pozycji w badaniu właściwym
<b>Rozwój organizacyjny</b>	<b>134</b>	<b>67</b>	<b>24</b>
Rozwój procesów	26	12	6
Rozwój struktur	31	12	4
Rozwój systemów operacyjnych	21	12	4
Rozwój systemów zarządzania	36	15	6
Rozwój kultury organizacyjnej	20	16	4
<b>Organizacyjne problemy wzrostu (konstrukt jednowymiarowy)</b>	<b>61</b>	<b>32</b>	<b>15</b>

Źródło: opracowanie własne

Tabela 2. Wyniki oceny dokonanej przez sędziów dla rozwoju struktur organizacyjnych (jednego z pięciu wymiarów rozwoju organizacyjnego)

Pytanie	Wszyscy sędziowie				Naukowcy											Praktycy								
	średnia	odch. stand.	średnia wewn.	ranking	średnia	odch. stand.	średnia wewn.	ranking	S1	S2	S3	S4	S5	S6	S1	S2	S3	S4	S5	S6	średnia	odch. stand.	średnia wewn.	ranking
1	4,17	0,94	4,30	4	4,33	0,82	4,50	5	4	4	5	5	3	5	4	5	4	2	5	4	4,00	1,10	4,25	7
2	3,42	1,31	3,50	20	2,83	1,33	3,00	23	2	4	4	1	4	2	3	5	3	5	3	4,00	1,10	4,00	13	
3	4,08	0,90	4,20	5	4,00	0,63	4,00	11	5	4	4	4	3	4	5	5	2	4	4	4,17	1,17	4,50	4	
4	3,50	1,24	3,60	16	3,83	1,33	4,00	11	5	3	2	5	3	5	4	1	3	4	3	3,17	1,17	3,50	22	
5	4,33	0,89	4,50	1	4,67	0,52	4,75	2	4	4	5	5	5	5	5	5	4	4	4	2	4,00	1,10	4,25	7
6	3,75	1,22	3,90	12	3,83	1,60	4,25	8	4	5	5	3	1	5	4	4	4	4	2	3,67	0,82	4,00	13	
7	3,42	1,44	3,50	20	2,83	1,83	2,75	27	1	5	2	3	1	5	4	3	4	4	5	4	4,00	0,63	4,00	13
8	3,42	1,38	3,50	20	3,33	1,63	3,50	16	4	5	2	3	1	5	4	4	2	4	5	2	3,50	1,22	3,50	22
9	3,17	0,94	3,10	27	3,00	1,10	2,75	27	3	3	2	3	5	2	3	4	2	3	4	4	3,33	0,82	3,50	22
10	3,58	1,38	3,70	14	3,17	1,60	3,25	20	2	5	3	3	1	5	5	4	2	4	5	4	4,00	1,10	4,25	7
11	4,08	0,51	4,10	7	4,00	0,63	4,00	11	4	4	4	3	5	4	4	4	4	4	5	4	4,17	0,41	4,00	13
12	3,42	0,79	3,40	22	2,83	0,41	3,00	23	3	3	3	3	3	2	5	3	4	4	4	4	4,00	0,63	4,00	13
13	3,42	1,38	3,50	20	2,33	0,82	2,50	30	1	3	2	3	3	2	3	3	5	5	4	4	4,50	0,84	4,75	2
14	3,25	1,06	3,20	25	3,17	1,33	3,00	23	2	4	2	4	5	2	3	2	4	4	4	4	3,33	0,82	3,50	22
15	3,08	0,90	3,10	27	3,17	0,98	3,25	20	2	4	2	4	3	4	3	4	2	4	2	2	3,00	0,89	3,00	29
16	3,00	1,13	3,00	29	2,67	1,37	2,50	30	2	3	2	1	5	3	4	3	2	3	4	4	3,33	0,82	3,50	22
17	4,17	1,34	4,40	3	3,83	1,83	4,25	8	2	5	5	5	1	5	5	4	5	4	5	4	4,50	0,55	4,50	4
18	4,25	1,06	4,40	2	4,17	0,98	4,25	8	3	4	5	3	5	5	5	4	5	4	5	2	4,33	1,21	4,75	2
19	4,00	1,35	4,20	6	4,00	1,67	4,50	5	1	5	5	5	3	5	4	4	5	4	5	2	4,00	1,10	4,25	7
20	3,50	1,38	3,60	16	3,33	1,63	3,50	16	1	5	2	5	3	4	3	5	4	5	2	3,67	1,21	3,75	17	
21	3,00	1,48	3,00	29	2,67	1,37	2,50	30	1	5	2	3	3	2	5	3	1	4	5	2	3,33	1,63	3,50	22
22	3,75	1,22	3,90	11	3,33	1,63	3,50	16	4	5	2	3	1	5	4	4	5	4	4	4	4,17	0,41	4,00	13
23	2,58	1,00	2,60	31	2,50	0,84	2,75	27	3	3	2	3	1	3	2	3	1	4	4	2	2,67	1,21	2,75	31
24	2,75	1,22	2,70	30	2,50	0,84	2,75	27	3	3	2	3	1	3	2	4	1	4	5	2	3,00	1,55	3,00	29
25	3,25	1,29	3,30	24	3,50	1,05	3,50	16	4	4	2	3	5	3	2	4	1	4	5	2	3,00	1,55	3,00	29
26	3,83	1,11	3,90	10	4,33	1,21	4,75	2	5	4	2	5	5	5	4	3	2	3	4	4	3,33	0,82	3,50	22
27	3,83	0,83	3,80	13	3,67	0,82	3,50	16	4	5	4	3	3	3	3	5	4	4	5	3	4,00	0,89	4,00	13
28	3,83	1,03	3,90	9	4,33	0,82	4,50	5	4	5	4	5	3	5	3	4	2	3	5	3	3,33	1,03	3,25	27
29	3,17	1,11	3,30	24	3,17	1,17	3,50	16	4	4	4	3	1	3	3	3	4	1	4	4	3,17	1,17	3,50	22
30	3,83	0,94	3,90	8	4,33	0,82	4,50	5	3	5	5	4	4	5	4	4	3	3	4	2	3,33	0,82	3,50	22
31	3,50	1,17	3,60	16	3,00	1,10	3,25	20	3	4	4	3	1	3	3	4	5	4	4	2	4,00	1,10	4,25	7

Źródło: opracowanie własne

stosunkiem głosów 11:1, że lepiej będzie zapytać respondentów wprost o tę dynamikę, niż pytać ich o poziom zmiennej obecnie i przed trzema laty<sup>6</sup>.

## Podsumowanie

Celem artykułu jest wzbogacenie dyskusji dotyczącej walidacji narzędzi pomiarowych w naukach o zarządzaniu przy wykorzystaniu opinii sędziów kompetentnych. W jego treści przedstawiono dylematy związane z metodyką pracy z sędziami przy selekcji wstępnych pozycji kwestionariuszowych do ankiety badań pilotażowych, następnie opisano procedurę zastosowaną przez autora przy opracowywaniu dwóch nowych skal pomiarowych.

Problemy, na jakie napotykają badacze, związane z brakiem jasnych wytycznych dotyczących kolejnych etapów pracy z ekspertami, a dotyczące przede wszystkim: liczby zaangażowanych w badanie sędziów, sposobu oceny przez nich wstępnych pozycji kwestionariusza oraz analizy otrzymanych odpowiedzi, sprawiają, że część badaczy może zrezygnować z tego kroku podczas opracowywania nowych skal. Na podstawie studiów literaturowych, dyskusji z ekspertami oraz doświadczeń nabytych podczas opracowywania własnych skal pomiarowych autor sugeruje zaangażowanie kilkunastu ekspertów będących w połowie teoretykami, w połowie zaś praktykami w danym obszarze. Warto polecić jest przeprowadzenie rozmów z kandydatami na sędziów przed podjęciem decyzji o ich zaangażowaniu w badanie, co da badaczowi możliwości: 1) przedstawienia przedmiotu badania, jego kontekstu oraz użytych definicji; 2) upewnienia się o posiadaniu przez kandydatów odpowiednich kompetencji w kontekście celu badania; 3) potwierdzenia chęci zaangażowania się w badanie. Do oceny przez sędziów przydatności poszczególnych stwierdzeń do pomiaru cechy ukrytej autor sugeruje wykorzystanie pięciostopniowej skali Likerta. Po otrzymaniu opinii sędziów zalecane jest oszacowanie stopnia zgodności ich ocen. W celu wyboru pozycji do kwestionariusza badań pilotażowych autor rekomenduje sporządzenie ich rankingu na podstawie średnich wewnętrznych ocen sędziów z pomocniczym użyciem wartości odchylenia standardowego oraz porównania średnich wewnętrznych policzonych w obu grupach sędziów.

Należy zdawać sobie sprawę z ograniczeń, którymi obarczone są opisane w publikacji badania. Po pierwsze, przegląd literatury, ze względu na cel artykułu oraz jego objętość, nie można uznać jako pełnego. Po drugie, wybór stwierdzeń do badania pilotażowego na podstawie średniej wewnętrznej może wywołać wątpliwość: czy pytanie z nieznacznie wyższą średnią wewnętrzną, jednak z wyższym odchyleniem standardowym, powinno zajmować wyższe miejsce w rankingu od pytania z nieznacznie niższą średnią, za to z niższym odchyleniem standardowym? Ograniczenia te wskazują jednocześnie na obszary prowadzenia dalszych analiz. Warto także wskazać obszary, które nie zostały poruszone w treści artykułu, mogące być przedmiotem dalszych rozważań. Nie można wykluczyć sytuacji, iż pomimo zachowania przez badacza odpowiedniej staranności i uzyskania od ekspertów zapewnienia o chęci wzięcia udziału

w badaniu, jeden lub kilku z nich oceni przedstawione stwierdzenia szybko, bez refleksji i wystarczającego zaangażowania. W przypadku przyznawania przez takiego eksperta jednakowych lub bardzo podobnych ocen wszystkim lub większości stwierdzeń, zauważenie tego faktu nie będzie stwarzać problemów. Nie można jednak wykluczyć sytuacji, iż niezaangażowany ekspert przyzna różne oceny poszczególnym stwierdzeniom. Czy możliwe jest uchwycenie niskiej jakości pracy eksperta w takiej sytuacji? Nawiązując do braku bezwzględnego obowiązku podążania za wynikami badania opinii sędziów, pojawić się może problem podważenia niewygodnej dla badacza opinii eksperta. Czy możliwe jest opracowanie katalogu przypadków, w których nieuwzględnienie opinii sędziego uznać należałoby za uzasadnione? Wydaje się także, że warto byłoby poznać opinie samych sędziów o trudnościach związanych z oceną. Ciekawe byłoby uchwycenie, czy występują różnice pomiędzy ocenami sędziów praktyków oraz sędziów naukowców. Czy można doszukać się tu jakichś prawidłowości? Jeśli tak, to z czego one wynikają?

Odpowiedzi na dylematy wskazane w artykule nie są wyczerpujące i jednoznaczne. Mając świadomość, że przedstawiona propozycja procedury pracy z sędziami nie jest jedyną, czy też najbardziej prawidłową procedurą, a przegląd literatury nie jest kompletny, autor ma nadzieję, że zaprezentowane w artykule rozważania i wyniki badań wpisują się w dyskusję dotyczącą opracowywania nowych narzędzi pomiarowych, będą stanowić zachętę dla badaczy do angażowania sędziów kompetentnych w proces opracowywania nowych skal pomiarowych z zachowaniem wysokiego stopnia rygoru metodologicznego oraz staną się przyczynkiem do dalszych pogłębianych analiz.

---

**dr hab. Maciej Czarnecki, prof. uczelni**  
**Uniwersytet Ekonomiczny we Wrocławiu**  
**Wydział Ekonomii i Finansów**  
**ORCID: 0000-0002-9694-8036**  
**e-mail: maciej.czarnecki@ue.wroc.pl**

## Przypisy

- 1) Skala pomiarowa, która nie została poddana walidacji, coś mierzy. Wbrew deklaracji, a często i przekonaniu badacza, najczęściej nie mierzy ona konstruktów, którymi powinna zmierzyć. Zaangażowanie sędziów kompetentnych w proces opracowywania nowej skali jest warunkiem koniecznym uzyskania przez nią walidacji treściowej oraz fasadowej (Czakoń, 2019, s. 5). Tematyka walidacji narzędzia pomiarowego jest dość obszerna i dobrze rozpoznana w literaturze (Churchill, 1979; 2002; DeVellis, 2017; Hornowska, 2017; Morgado i in., 2018; Czakoń, 2019). Jej omawianie wykracza poza ramy niniejszego artykułu.
- 2) Pojęcie *pozycja kwestionariuszowa* jest używane w literaturze przedmiotu zamiennie z pojęciami: pytanie, stwierdzenie oraz pozycja ankiety.
- 3) Wskaźnik ten dotyczy *utruty w pewności przypisywanej złym decyzjom sędziów* w wyniki braku zgodności między nimi i porównywane jest przez R.T. Rust i B. Cooil (1994, s. 4) do wskaźnika Alfa Cronbacha.

- 4) Dziękuję za udział w dyskusji i udzieloną pomoc: prof. dr hab. M. Rószkiewicz, dr hab. A. Austen, drowi hab. P. Zbierowskiemu, drowi hab. Ł. Kuźmińskiemu, drowi hab. W. Grabowskiemu oraz dr C. Kozyrze.
- 5) Niskie oceny uzyskały m.in. takie stwierdzenia, jak: „Inwestujemy w technologie umożliwiające pozyskiwanie w czasie rzeczywistym informacji o naszych klientach i ich potrzebach”, „Mamy w firmie wdrożony dedykowany system CRM (Customer Relationship Management)” (Wang, Feng, 2012, s. 124), „Większość ludzi uważa, że narady firmowe są stratą czasu”, „Ludzie uważają, że ciągle brakuje czasu na realizację zadań” (Flamholtz, Randle, 2012, s. 49) itd.
- 6) Wszyscy sędziowie uzasadnili swoje wybory. Wśród uzasadnień większość miała charakter zwięzły (np. „metoda pytania wprost o dynamikę lepiej precyzuje zmiany w nasileniu zjawisk”), kilka z nich przyjęło formę bardziej rozbudowanych rozważań.

## Bibliografia

- [1] Cabała P. (2010), *Zastosowanie współczynnika konkordancji w pomiarze zgodności ocen ekspertów*, „Przegląd Statystyczny”, Nr 57(2–3), s. 36–52.
- [2] Churchill G.A. (1979), *A Paradigm for Developing Better Measures of Marketing Constructs*, „Journal of Marketing Research”, Vol. 16, No. 1, pp. 64–73.
- [3] Churchill G.A. (2002), *Badania marketingowe: podstawy metodologiczne*, Wydawnictwo Naukowe PWN, Warszawa
- [4] Crawford J.A., Kelder J.A. (2019), *Do We Measure Leadership Effectively? Articulating and Evaluating Scale Development Psychometrics for Best Practice*, „The Leadership Quarterly”, Vol. 30, No. 1, pp. 133–144.
- [5] Czakon W. (2019), *Walidacja narzędzia pomiarowego w naukach o zarządzaniu*, „Przegląd Organizacji”, Nr 4, s. 3–10.
- [6] Czarnecki M. (2016), *Rozwój organizacyjny a problemy wzrostu przedsiębiorstw*, Wydawnictwo Difin, Warszawa.
- [7] Danneels E. (2016), *Survey Measures of First-and Second-order Competences*, „Strategic Management Journal”, Vol. 37, No. 10, pp. 2174–2188.
- [8] DeVellis R.F. (2017), *Scale Development: Theory and Applications*, Sage Publications, Los Angeles.
- [9] Flamholtz E.G., Randle Y. (2012), *Growing Pains: Transitioning from an Entrepreneurship to a Professionally Managed Firm*, John Wiley & Sons, New York.
- [10] Gorbaniuk O. (2016), *Wykorzystywanie procedury sędziów kompetentnych w naukach społecznych i możliwości jej oceny psychometrycznej za pomocą narzędzi dostępnych w Statistica*, [w:] J. Jakubowski, J. Wątroba (red.), *Zastosowanie statystyki i data mining w badaniach naukowych*, Statsoft Polska, Kraków, s. 5–20.
- [11] Govindarajan V., Kopalle P.K. (2006), *Disruptiveness of Innovations: Measurement and an Assessment of Reliability and Validity*, „Strategic Management Journal”, Vol. 27, No. 2, pp. 189–199.
- [12] Hardesty D.M., Bearden W.O. (2004), *The Use of Expert Judges in Scale Development: Implications for Improving Face Validity of Measures of Unobservable Constructs*, „Journal of Business Research”, Vol. 57, No. 2, pp. 98–107.
- [13] Hornowska E. (2017), *Testy psychologiczne: teoria i praktyka*, Wydawnictwo Naukowe Scholar, Warszawa
- [14] Lavrakas P.J. (2008), *Encyclopedia of Survey Research Methods*, Sage Publications, Los Angeles.
- [15] Liang J., Farh C.I., Farh J.L. (2012), *Psychological Antecedents of Promotive and Prohibitive Voice: A Two-wave Examination*, „Academy of Management Journal”, Vol. 55, No. 1, pp. 71–92.
- [16] Morgado F.F., Meireles J.F., Neves C.M., Amaral A.C., Ferreira M.E. (2018), *Scale Development: Ten Main Limitations and Recommendations to Improve Future Research Practices*, „Psicologia: Reflexão e Crítica”, Vol. 30, No. 1, pp. 1–20.
- [17] Obermiller C., Spangenberg E.R. (1988), *Development of a Scale to Measure Consumer Skepticism toward Advertising*, „Journal of Consumer Psychology”, Vol. 7, No. 2, pp. 159–186.
- [18] Rai R.K. (2016), *A Co-opetition-based Approach to Value Creation in Interfirm Alliances: Construction of a Measure and Examination of Its Psychometric Properties*, „Journal of Management”, Vol. 42, No. 6, pp. 1663–1699.
- [19] Rust R.T., Cooil B. (1994), *Reliability Measures for Qualitative Data: Theory and Implications*, „Journal of Marketing Research”, Vol. 31, No. 1, s. 1–14.
- [20] Thomas B., Lucas K. (2019), *Development and Validation of the Workplace Dignity Scale*, „Group & Organization Management”, Vol. 44, No. 1, pp. 72–111.
- [21] Wang Y., Feng H. (2012), *Customer Relationship Management Capabilities. Measurement, Antecedents and Consequences*, „Management Decision”, Vol. 50, No. 1, pp. 115–129.
- [22] Zaichkowsky J.L. (1985), *Measuring the Involvement Construct*, „Journal of Consumer Research”, Vol. 12, No. 3, pp. 341–352.

## Working with Expert Judges in the Process of New Measure Scale Development: Methodological Guidelines and Case Study

### Summary

One of the goals of management science is finding relations between organisational phenomena. Those are not directly observable, thus a need for their operationalisation and developing appropriate measure scales arises. According to the state of the art, external experts should be involved in this process (a.k.a expert judges). However, there is a lack of methodological guidelines in the literature concerning that issue. The dilemmas concern, among others, the number of experts involved, methods of their assessment of initial questionnaire items and analysis of the judges' responses. The article presents the Author's recommendations regarding the indicated dilemmas, along with the exemplification. They are based on a literature review, discussions with management science community representatives and experts in the field of statistics. The Author proposes the involvement of 12 judges: 6 scholars and 6 practitioners using a five-point Likert scale to assess the initial questionnaire items. He suggests their responses' assessment with the use of internal mean with the auxiliary use of the standard deviation and a comparison of the means calculated in groups of judges.

### Keywords

judge experts, questionnaire, measure scale, measure tool